

IZA DP No. 4364

**The Research Potential of New Types of
Enterprise Data Based on Surveys from
Official Statistics in Germany**

Joachim Wagner

August 2009

The Research Potential of New Types of Enterprise Data Based on Surveys from Official Statistics in Germany

Joachim Wagner
*Leuphana University Lueneburg
and IZA*

Discussion Paper No. 4364
August 2009

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

The Research Potential of New Types of Enterprise Data Based on Surveys from Official Statistics in Germany^{*}

A new generation of data sets became available recently in the research data centres of the German statistical offices. These new data combine information for firms gathered in different surveys (or from other sources) that could not be analyzed jointly before. This paper offers a short description of these data, and gives examples of their use to demonstrate their research potential. Furthermore, and looking ahead to the next generation of data, it discusses an ongoing project, *KombiFiD*, that will for the first time offer access to linked firm level data collected by different data producing public institutions.

JEL Classification: C8

Keywords: enterprise data, Germany, research data centres

Corresponding author:

Joachim Wagner
Leuphana University Lueneburg
Institute of Economics
P.O.Box 2440
D-21314 Lueneburg
Germany
E-mail: wagner@leuphana.de

^{*} Thanks to Nils Braakmann and Alexander Vogel for critical remarks on an earlier version.

1. Introduction

When the Federal Statistical Office and the statistical offices of the German federal states opened research data centres (described in detail in Zühlke et al. 2004) in 2001 and 2002 a new era started for researchers working in empirical economics. Access to confidential data for individuals and firms that were collected in surveys performed by the statistical offices was no longer impossible or extremely difficult. From now on researchers could easily (and at low costs) use the research data centres (RDC) as a gateway to the nuggets they were digging for. The number and variety of data sets provided by the RDC increased steadily (see Kaiser and Wagner (2008) for an overview), and so did the use of it by researchers. The high potential of these data as a basis to generate new stylized facts, to motivate assumptions used in formal theoretical models, to test theoretical hypotheses econometrically, and to be used in policy consultation and evaluation, is documented in a large and growing number of publications.¹

From their start the RDC not only offered access to micro level data from official statistics as cross-section data from one survey. Panel data sets that linked results from various waves of a survey over time were an innovative product provided by the RDC.² This type of data enormously extended the research potential of data from official statistics by allowing dynamic analyses and control for unobserved heterogeneity via panel econometric methods.

Compared to this first generation of firm panel data sets, a second generation of data sets that became available only recently has an even higher research potential. These new data combine information for firms³ gathered in different surveys (or from external sources) that could not be analyzed jointly before. This paper offers a short description of these data, and gives examples of their use to demonstrate their research potential. Furthermore, and looking ahead to the next generation of data, it discusses an ongoing project, *KombiFiD*, that will for the first time offer access to linked firm level data collected by different data producing public institutions.

¹ For partial surveys, see Wagner (2007a, 2008a).

² It should be noted that panel data constructed from surveys conducted by statistical offices in Germany had been used for many years before, mostly by researchers active in the FiDASt network (see Wagner 2008b). While these data were limited to one federal state at a time, the RDC for the first time provided panel data for Germany as whole, linking surveys across time and space.

³ Note that due to legal rules this type of matched data is only available for firms, not for individuals.

2. *AFiD* - Combining data from various regular surveys

Data collected in surveys from official statistics have, among others, two big advantages compared to data from surveys administered by other institutions – the units covered by the survey are usually obliged to report (and to report the true figures), and the survey often is a census covering all units from a well-defined population. Therefore, data from official statistics are high quality data. However, there is one big disadvantage that limits the usefulness of these data for scientific analyses – the information collected in a survey is limited to a few variables with a focus on one topic (or a limited number of topics).

A case in point is the monthly report of establishments from mining and manufacturing. Details aside all establishments from these industries (that have more than the required minimum number of employees) have to report the number of employees, hours worked, wages paid, sales in Germany, and exports (see Konold 2007). While these data are an important basis (and have been extensively used) for empirical investigations of exports (see, e.g. Wagner 2007b), the use of the data for this purpose is severely limited by the absence of any information on important determinants of exports (e.g., productivity measured as value added per employee, or research and development activities) and on a core indicator of firm performance, profitability. Information on value added per employee, research and development activities, and profitability, however, is either collected in a different survey, the cost structure survey (see Fritsch et al. 2004), or can be computed from these data (which do not contain any information on exports). If micro data from these two surveys are merged, the research potential of both surveys is increased considerably.

Merging firm level data from different surveys to construct data sets that cover information on a wider range of variables than the ones collected in any of these surveys, one at a time, is the basic idea of the project *AFiD*. *AFiD* is an acronym for the German *Amtliche Firmendaten für Deutschland* (official firm data for Germany). Merging of firm data from different sources of official statistics is legal according to §13a BStatG (Bundesstatistikgesetz, or federal statistics law), and it is technically feasible because an identical firm identifier is used in the different surveys.

In the *AFiD* project, which is in detail described in Malchin and Voshage (2009), several different panel data sets are provided in the RDC, including the *AFiD-Panels Agriculture, Industrial Units, Industrial Enterprises, Energy Units, Services,*

and *Business Register*. For some of these panels the information potential can be enlarged even further by adding variables from the so-called *AFID-Modules* Earnings, Use of Energy, and the environmental moduls.⁴

This new type of combined data is an important innovation.⁵ To illustrate the research potential, two recent studies based on such combined data are summarized here:

The first example is a study by Fryges and Wagner (2009) that deals with the causal effect of exports and profits. Their starting point is the stylized fact⁶ that exporting firms are more productive than otherwise identical firms that sell on the national market only. Exporting firms have to bear extra costs due to, among others, market research, adaptation of products to local regulations, or transport costs. These extra costs are one reason for a self-selection of the more productive firms on international markets. Furthermore, exporting firms tend to pay higher wages than non-exporting firms. A question that has not been investigated in this literature is whether the productivity advantage of exporting firms does lead to a profitability advantage of exporters compared to otherwise identical non-exporters even when exporters are facing extra costs and pay higher wages.

Profitability is a performance dimension that has not been investigated empirically in the literature dealing with the micro-econometrics of international trade. The reason for this missing evidence is that information on profits is no included in the firm level data sets used to investigate exports. Fryges and Wagner (2009) use a data set of the *AFiD* panel type that matches data on exports from the monthly survey of establishments in mining and manufacturing (mentioned above), aggregated at the enterprise level and over the month of a year, with data on profitability and other firm characteristics from the cost structure survey (mentioned above, too). They document a positive profitability differential of exporters compared to non-exporters that is statistically significant, though rather small, when observed firm characteristics and unobserved firm specific effects are controlled for. In contrast to nearly all empirical studies on the relationship between productivity and exports

⁴ Note that tailor made variants that combine data from various surveys according to a wish-list provided by a researcher can be prepared on request.

⁵ It should be noted, however, that panel data combined from the monthly report of establishments in mining and manufacturing, from the annual investment survey, and from the annual survey of small establishments have been used for a long time in the *FiDASt* network mentioned above. These combined data that can be viewed as a role model for the *AFiD* panels are described in Wagner (2000).

⁶ See Fryges and Wagner (2009) for references to the literature on this and other points mentioned here.

they do not find any evidence for self-selection of more profitable firms into export markets. Furthermore, they use a newly developed continuous treatment approach and show that exporting improves the profitability almost over the whole range of the export-sales ratio. This means, that the usually observed higher productivity of exporters is not completely absorbed by the extra costs of exporting or by higher wages paid by internationally active firms. This evidence for Germany, a leading actor on the world market for manufactured goods, is interesting on its own, and it can serve as a benchmark for future studies using comparable data from other countries.

The second example for the use of AFID panel type data is a study by Wagner (2009) dealing with the extent and causes of product diversification in German manufacturing enterprises. Interest in the theoretical and empirical analysis of multi-product enterprises is growing. This paper uses representative longitudinal data from the production survey covering the years 1995 to 2004 to document for the first time the extent of product diversification in German manufacturing enterprises and the role of single- and multi-product firms. In 2004 about 60 percent of all enterprises were multi-product firms, producing 4.4 products on average. Multi-product firms are more often found among the larger enterprises. In 2004 the share of multi-product enterprises in total sales and total exports was 81 percent and 84 percent, respectively. Enterprises producing a large number of products are a rare species – only about three percent of all firms produce more than ten different products. This extent of product diversification is rather stable over the time period under consideration.

Based on combined data from the production survey and data from the cost structure survey (mentioned above) the empirical investigation demonstrates that compared to single-product firms multi-product firms are larger, have a higher productivity, and both a higher human capital intensity and research and development intensity. These characteristics of multi-product firms do exist before single-product enterprises turn to multi-product enterprises. Therefore, and in accordance with the *resource view*, they can be considered as determinants of product diversification at the firm level.

Both examples illustrate that the research potential of panel data matched from various surveys is much larger than the research potential of the data from only one survey.

3. *AFiDextended* – Combining data from regular surveys and special purpose surveys

Besides the regular surveys with mandatory participation of the units in the population (or the sample) from time to time the statistical offices perform so-called special purpose surveys (*Erhebungen für besondere Zwecke*, see §7 of the federal statistics law BStatG). Participation in these special purpose surveys is voluntary, and the sample is limited to 20.000 units. A prerequisite for this kind of survey is either a pressing need for data in the process of preparing or substantiating a planned decision by a high government agency, or the clarification of a methodological question in statistics.

Information collected in such a special purpose survey is by definition not available from other sources, at least not in comparable quality and quantity. Data from these surveys, therefore, can be an interesting source for econometric investigations. A case in point is the survey on relocation of economic activities (*Verlagerung wirtschaftlicher Aktivitäten*) conducted in 2006 (see Zwania (2008)). The enterprises were asked about the reasons to relocate production inside Germany and across the German border, the role of barriers to relocate, the extent of relocation in the past and plans for the near future, the regions they relocated to, the partners in the relocation process, and the consequences of relocation for the firm. For the first time information on these topics is available for a large sample of firms from a survey performed by official statistics, and descriptive results offer a number of new interesting facts on these important (and in part highly controversial) topics (see Statistisches Bundesamt (2008)).

However, the research potential of the data from the relocation survey as such is limited. First of all, it is a cross section survey only, and this hinders any dynamic or causal analyses. Second, many questions ask for a subjective assessment by the interviewee, and it is an open question whether this persons is willing and able to give a correct answer. To enhance the research potential of these data considerably, they were matched to an *AFiD* type panel data set that has information from the monthly report and from the cost structure survey (both mentioned above).⁷ This is one example of a new type of data that I label *AFiDextended*.

⁷ Matching is technically feasible by using the enterprise number from the special purpose survey that is identical to the enterprise number used in regular surveys, and it is legal according to §13a BStatG.

The combined data from matched regular surveys and the special purpose survey on relocation are used in a number of ongoing econometric investigations. One of these studies asks for the causal effect of relocation on various dimensions of firm performance (see Wagner (2009c)). Data from the relocation survey were used to identify enterprises that relocated production abroad in the period 2001-03 for the first time. These firms are compared to firms that did not relocate production abroad between 2000 and 2006. The comparison is based on data from the monthly report and the cost structure survey, and it is performed for both 2004 (to document differences between the two groups of firms after some of them started to relocate abroad) and for 2000 (when none of them did relocate abroad). It turns out that compared to non-relocating firm relocating firms are larger and more productive, more human capital intensive, have a higher share of exports in total sales, a higher share of employees working in research and development, and use more capital per employee (while the rate of profit does not differ between the two groups). All these differences existed in 2000, the year before some firms started to relocate, and this points to self-selection of “better” firms into offshoring.

To investigate the causal effects of relocation across borders on firm performance, matched pairs of firms that did and did not start to relocate abroad in 2001-03 were formed based on firm characteristics in 2000 and their development between 1997 and 2000, and the performance of both groups was compared for 2004-06 when one half of the twin firms was made of relocating firms and the other half was not. While the performance of offshoring firms was on average better than that of non-relocating firms (with the exception of the growth of capital intensity) these differences, however, were statistically significant from zero only in the case of the growth of productivity. To state it differently, contrary to what is often argued we find no evidence for a negative causal effect of offshoring on employment or other core dimensions of firm performance.

This example illustrates that the research potential of data from a one-time cross section special purpose survey can be enhanced considerably by matching the data to panel data already available from regular surveys to build an *AFiDextended* data set.

4. *AFiDplus* – Combining data from official statistics and external sources

Surveys from official statistics do not cover each and every characteristic of a firm researchers are interested in. A case in point is information on codetermination at the enterprise level (i.e., *Unternehmensmitbestimmung*) where employees are sitting on supervisory boards. Details aside, there is one-third codetermination in enterprises with between 500 and 2,000 employees according to the 2004 Third Party Act (*Drittelbeteiligungsgesetz*). However, experts for labour law and industrial relations point out that, contrary to the law, limited liability companies (*Gesellschaften mit beschränkter Haftung, GmbH*) from this size class very often do not have a supervisory board at all, and, therefore, are not co-determined firms. No information on this is collected in surveys from official statistics.

Such information, however, is available from a commercial data base, the *Hoppenstedt Datenbank Grossunternehmen* (see www.hoppenstedt-grossunternehmen.de). This data base contains information on the 25,000 largest enterprises in Germany with at least 200 employees and/or a sales volume of at least 20 Mio. €. Coverage for enterprises with 500 or more employees is complete in this data base. In this data base it is reported whether or not an enterprise has a supervisory board (and its size), and whether or not worker representatives are among the board members (and their number). This information on the presence of a supervisory board and its composition, however, is not available for all enterprises. As a part of an ongoing project Troch (2009) collected missing information for limited liability enterprises from manufacturing industries in West Germany via telephone calls.⁸

The most important result from an empirical investigation using these data is that only some 60 percent of all limited liability companies from German manufacturing industries that fell under the Third Party Act of 2004 had a co-determined supervisory board in 2007/2008 (Troch 2009). This provides evidence for the existence of a large co-determination free zone among limited liability companies with 500 to 2.000 employees.

This fact offers the possibility for an empirical investigation of the relationship between supervisory board level co-determination and firm performance based on a

⁸ A detailed description of the data and the process of its collection can be found in Troch (2009). Boneberg (2009) reports comparable data for the West German services sector industries.

direct comparison of co-determined and co-determination free firms from the same size class with the same legal form. However, information on firm performance (productivity, defined as value added per employee, and profitability, measured by gross firm surplus divided by total sales) is missing in the Hoppenstedt data base. Therefore, to perform this investigation, information on the presence of a co-determined supervisory board in an enterprise (taken from the Hoppenstedt data base, or collected via telephone) was merged with data from a second source, the cost structure survey for enterprises in the manufacturing (mentioned above). This is an example for a new type of data that I label *AFiDplus*.

Merging was done using information about the register number and register court of the trade register (*Handelsregisternummer* und *Handelsregistergericht*) for the enterprise, because this information is available in both the Hoppenstedt data base and in the official register of enterprises (*Unternehmensregister*) that was linked with the cost structure survey data. Note that merging firm level data from official statistics and from other sources is legal according to §13a of the federal statistics law (*Bundesstatistikgesetz*) provided the data from external sources are publicly available. This is the case with the data on codetermination used here, because they are either taken from the publicly (though not costless) available Hoppenstedt data base, or published on the web as an appendix to Troch (2009).

Using this new type of data that combines information on the codetermination status of enterprises from a commercial data base and supplementary information collected from the firms with comprehensive data on the firms from official statistics a study by Wagner (2009b) contributes to the small empirical literature on the co-determination – firm performance nexus. The data allow for the first time a direct comparison of enterprises from the same size class with and without co-determination at the supervisory board level. It is shown that one-third codetermination at the supervisory board level in limited-liability companies from West German manufacturing industries seems to be neither positively nor negatively related to two core firm performance indicators, productivity and profitability.

This example illustrates the high research potential of *AFiDplus* type data that combine panel data already available from regular surveys with data from an external source that is not affiliated with official statistics. To pursue this strategy it must be technically feasible to match these external data to *AFiD* type panel data, and this is most easily done via the register number and register court of the trade register,

because this information is readily available in both *AFiD* type panel data and enterprise data from commercial sources.⁹ Furthermore, the external data have to be publicly available (but not necessarily for free) for the matching to be legal, and this condition can easily be fulfilled for newly collected data by publishing it on the web. That said, it can be expected that more and more researchers will use *AFiDplus* type data in the future.

5. A look ahead: *KombiFiD* – Combining data across institutions

The first and second generation of panel data sets available at the RDC of the German statistical offices that were discussed in the earlier sections of this paper are all based on confidential survey data collected by official statistics. As is well known, there are other official institutions besides the statistical offices that are involved in the collection and preparation of confidential micro data, and some of these data are firm level data. Cases in point are data on foreign direct investments collected by the German Central Bank (*Deutsche Bundesbank*) described in Lipponer (2003), and detailed data on the structure of the employees in the firms prepared by the Federal Employment Agency (*Bundesagentur für Arbeit*) on the basis of social security records (see Spengler 2008).

These confidential firm level data that are kept strictly inside the data producing institutions (but can be accessed by researchers via the RDC of these institutions) are complements – often information that is in one data set is missing in all other data sets. The high potential for research that would be offered by data sets that combine firm level data provided by different data producing institutions is obvious. To give just one example, up to now there is no large scale high-quality firm level data set that has information on exports, imports, and foreign direct investments. If data from official statistics on exports and imports could be matched with data on foreign direct investments from the Bundesbank it would be possible to investigate the relation between these different forms of international firm activities.

However, in Germany merging of confidential firm level data across the boundaries of the data producers is still in its infancy. One reason is that it is sometimes technically not trivial, because the firm identifiers used by the data

⁹ If it is not, it can easily be added to any self-made data set as long as the name of the enterprise is known because information on the register number and register court of the trade register is available online from www.handelsregister.de.

producers are not identical. The more important obstacle is that according to the law cross-institutional merging of firm level data is only allowed if the firms gave their written consent to do so. In an ongoing project, *KombiFiD*,¹⁰ a pilot study is carried out that asks a large number of firms for their consent to merge data from the statistical offices, the Federal Employment Agency, and the Bundesbank (see Bender, Wagner and Zwick (2007) and Braakmann (2009)). The combined data will then be available for researchers in the RDC of the data producers to explore the research potential of this next generation of firm panel data.¹¹

References

- Bender, Stefan, Joachim Wagner and Markus Zwick (2007), *KombiFid – Kombinierte Firmendaten für Deutschland*. University of Lüneburg Working Paper Series in Economics No. 60, September.
- Boneberg, Franziska (2009), Die Drittelmitbestimmungslücke im Dienstleistungssektor: Ausmaß und Bestimmungsgründe. University of Lüneburg Working Paper Series in Economics No. 114, January.
- Braakmann, Nils (2009), “Combined firm data for Germany” – The concept, a comparison with selected data sets and an assessment of its research potential. Unpublished, March.
- Fritsch, Michael, Bernd Görzig, Ottmar Hennchen and Andreas Stephan (2004), Cost Structure Surveys in Germany. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 124(4), 557-566.
- Fryges, Helmut and Joachim Wagner (2009), Exports and Profitability – First Evidence for German Manufacturing Firms. *The World Economy* 32 (in Press).
- Kaiser, Ulrich and Joachim Wagner (2008), Neue Möglichkeiten zur Nutzung vertraulicher amtlicher Personen- und Firmendaten. *Perspektiven der Wirtschaftspolitik* 9(3), 329-349.

¹⁰ This acronym stands for *Kombinierte Firmendaten für Deutschland* (combined firm level data for Germany).

¹¹ Up-to-date information on the KombiFiD project and how to access the data can be found on www.kombifid.de.

- Konold, Michael (2007), New possibilities for economic research through integration of establishment-level panel data of German Official Statistics. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 127(2), 321-334.
- Lipponer, Alexander (2003), Deutsche Bundesbank's FDI micro database. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 123(4), 593-600.
- Malchin, Anja and Ramona Voshage (2009), Official Firm Data for Germany. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 129(3) (in press).
- Spengler, Anja (2008), The Establishment History Panel. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 128(3), 501-509.
- Statistisches Bundesamt (2008), Verlagerung wirtschaftlicher Aktivitäten. Ergebnisse der Piloterhebung. Wiesbaden: Statistisches Bundesamt, August.
- Troch, Sebastian (2009), Drittelbeteiligung im Aufsichtsrat – Gesetzliche Regelung versus Unternehmenspraxis. Ausmaß und Bestimmungsgründe der Umgehung des Drittelbeteiligungsgesetzes in Industrieunternehmen. University of Lüneburg Working Paper Series in Economics No. 128, May.
- Wagner, Joachim (2000), Firm Panel Data from Official Statistics. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 120(1), 143-150.
- Wagner, Joachim (2007a), Politikrelevante Folgerungen aus Analysen mit Firmendaten der amtlichen Statistik. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 126(3), 359-374.
- Wagner, Joachim (2007b), Exports and Productivity in Germany. *Applied Economics Quarterly* 53(4), 353-373.
- Wagner, Joachim (2008a), Die Forschungspotenziale der Betriebspaneldaten des Monatsberichts im Verarbeitenden Gewerbe. *AStA – Wirtschafts- und Sozialstatistisches Archiv* 2(2), 209-221.
- Wagner, Joachim (2008b), FiDASt Reloaded. In: Rolf, Gabriele, Markus Zwick and Gert G. Wagner (Eds.), *Fortschritte in der informationellen Infrastruktur in Deutschland*. Baden-Baden: Nomos, 369-380.
- Wagner, Joachim (2009a), Produktdifferenzierung in deutschen Industrieunternehmen 1995 – 2004: Ausmaß und Bestimmungsgründe. *Jahrbücher für Nationalökonomie und Statistik* 229 (im Druck).

- Wagner, Joachim (2009b), One-third codetermination at company supervisory boards and firm performance in German manufacturing industries: First direct evidence from a new type of enterprise data. University of Lueneburg Working Papers in Economics 134, August.
- Wagner, Joachim (2009c), The causal effects of offshoring on firm performance: First evidence for German manufacturing enterprises. Unpublished draft paper, August.
- Zühlke, Sylvia, Markus Zwick, Sebastian Scharnhorst and Thomas Wende (2004), The research data centres of the Federal Statistical Office and the statistical offices of the *Länder*. *Schmollers Jahrbuch / Journal of Applied Social Science Studies* 124(4), 567-578.
- Zwania, Jacek (2008), Verlagerung wirtschaftlicher Aktivitäten. Theoretischer Hintergrund zur Erhebung. *Wirtschaft und Statistik* 6, 477-482.